Exploratory Data Analysis

Lecture 9

Corina Besliu

Technical University of Moldova

September 28, 2021

TECHNICAL UNIVERSITY
OF MOLDOVA

What is Hypothesis Testing

To understand hypothesis testing let us start with nonstatistical applications of it, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case, and a jury must make a decision on the basis of the evidence presented.

In fact, the jury conducts a <u>test of hypothesis</u>.

- The first is called the null hypothesis and is represented by H_0:
  H_0: The defendant is innocent.
- The second is called the alternative hypothesis and is denoted H_A:
  H_A: The defendant is guilty.

What is Hypothesis Testing

To understand hypothesis testing let us start with nonstatistical applications of it, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case, and a jury must make a decision on the basis of the evidence presented.

In fact, the jury conducts a test of hypothesis.

- The first is called the null hypothesis and is represented by H_0:
  H_0: The defendant is innocent.
- The second is called the alternative hypothesis and is denoted H_A:
  H_A: The defendant is guilty.

What is Hypothesis Testing

To understand hypothesis testing let us start with nonstatistical applications of it, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case, and a jury must make a decision on the basis of the evidence presented.

In fact, the jury conducts a test of hypothesis.

- The first is called the null hypothesis and is represented by H_0:
  H_0: The defendant is innocent.
- The second is called the alternative hypothesis and is denoted H_A:
  H_A: The defendant is guilty.

What is Hypothesis Testing

To understand hypothesis testing let us start with nonstatistical applications of it, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case, and a jury must make a decision on the basis of the evidence presented.

In fact, the jury conducts a test of hypothesis.

- The first is called the null hypothesis and is represented by H_0:

  H_0: The defendant is innocent.

- The second is called the alternative hypothesis and is denoted H_A:

  H_A: The defendant is guilty.

What is Hypothesis Testing

To understand hypothesis testing let us start with nonstatistical applications of it, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case, and a jury must make a decision on the basis of the evidence presented.

In fact, the jury conducts a test of hypothesis.

- The first is called the null hypothesis and is represented by H_0:
  H_0: The defendant is innocent.
- The second is called the alternative hypothesis and is denoted H_A:
  H_A: The defendant is guilty.

What is Hypothesis Testing

To understand hypothesis testing let us start with nonstatistical applications of it, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case, and a jury must make a decision on the basis of the evidence presented.

In fact, the jury conducts a test of hypothesis.

- The first is called the null hypothesis and is represented by H_0:
  H_0: The defendant is innocent.
- The second is called the alternative hypothesis and is denoted H_A:
  H_A: The defendant is guilty.

What is Hypothesis Testing

To understand hypothesis testing let us start with nonstatistical applications of it, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case, and a jury must make a decision on the basis of the evidence presented.

In fact, the jury conducts a test of hypothesis.

- The first is called the null hypothesis and is represented by H_0:
  H_0: The defendant is innocent.
- The second is called the alternative hypothesis and is denoted H_A:
  H_A: The defendant is guilty.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
  - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
  - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
    - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
    - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
  - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
  - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
  - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
  - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
    - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
    - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
    - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
    - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
  - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
  - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
  - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
  - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

What is Hypothesis Testing

There are only two possible decisions:

- Convicting the defendant $\rightarrow$ rejecting the null hypothesis (H_0) in favor of the alternative H_A.
    - There was enough evidence to conclude that the defendant was guilty.
- Acquitting the defendant $\rightarrow$ not rejecting the null hypothesis (H_0) in favor of the alternative H_A
    - There was not enough evidence to conclude that the defendant was guilty.

Notice that we do not say that we accept the null hypothesis!

This would be equal to saying that the defendant is innocent, and this is not a decision that the justice system is allowed to make.

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error → reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error → not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related → any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
    - An innocent person is wrongly convicted.
    - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
    - A guilty defendant is acquitted.
    - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious.
Interestingly in statistics it is also the case (most of the times).

Type I error & Type II error

There are two possible errors:

- Type I error $\rightarrow$ reject a true H_0.
  - An innocent person is wrongly convicted.
  - The probability of a Type I error $= \alpha$ (also called the *significance level*)
- Type II error $\rightarrow$ not rejecting a false H_0.
  - A guilty defendant is acquitted.
  - The probability of a Type II error $= \beta$

The error probabilities α and β are inversely related $\rightarrow$ any attempt to reduce one will increase the other.

In 'humane' justice systems Type I errors are regarded as more serious. Interestingly in statistics it is also the case (most of the times).

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $= \beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $= \beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $=\beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $=\beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
    - the null hypothesis (H_0).
    - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that $\underline{H_0 \text{ is not true}}$.
4. There are two possible decisions:
    - Conclude that there is enough evidence to reject H_0
    - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
    - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
    - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $=\beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $=\beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $=\beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that $\underline{H_0 \text{ is not true}}$.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $= \beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $= \beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that H_0 is not true.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error → reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error → don't reject a false H_0 (P(Type I error) $= \beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that $\underline{H_0 \text{ is not true}}$.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $=\beta$).

Lets extend these concepts to statistical hypothesis testing

Summary of Hypothesis Testing

1. There are two hypotheses:
   - the null hypothesis (H_0).
   - the alternative hypothesis (H_A).
2. The testing procedure begins with the assumption that the H_0 is true.
3. The goal of the process is to determine whether there is enough evidence to infer that <u>H_0 is not true</u>.
4. There are two possible decisions:
   - Conclude that there is enough evidence to reject H_0
   - Conclude that there is not enough evidence to reject H_0
5. Two possible errors can be made in any test:
   - Type I error $\rightarrow$ reject a true H_0 (P(Type I error) $= \alpha$).
   - Type II error $\rightarrow$ don't reject a false H_0 (P(Type I error) $=\beta$).

Lets extend these concepts to statistical hypothesis testing

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support H_A $\rightarrow$ reject H_0 $\rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow$ X is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support H_A $\rightarrow$ **cannot reject** H_0 $\rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow$ X is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0\colon \beta = 0$$
$$H_A\colon \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support $H_A \rightarrow$ reject $H_0 \rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow X$ is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support $H_A \rightarrow$ **cannot reject** $H_0 \rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow X$ is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0\text{: } \beta = 0$$
$$H_A\text{: } \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support H_A $\rightarrow$ reject H_0 $\rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow$ X is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support H_A $\rightarrow$ **cannot reject** H_0 $\rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow$ X is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support $H_A \rightarrow$ reject $H_0 \rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow X$ is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support $H_A \rightarrow$ **cannot reject** $H_0 \rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow X$ is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support H_A $\rightarrow$ reject H_0 $\rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow$ X is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support H_A $\rightarrow$ **cannot reject** H_0 $\rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow$ X is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support H_A $\rightarrow$ reject H_0 $\rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow$ X is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support H_A $\rightarrow$ **cannot reject** H_0 $\rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow$ X is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support H_A $\rightarrow$ reject H_0 $\rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow$ X is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support H_A $\rightarrow$ **cannot reject** H_0 $\rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow$ X is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support $H_A \rightarrow$ reject $H_0 \rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow X$ is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support $H_A \rightarrow$ **cannot reject** $H_0 \rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow X$ is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support H_A $\rightarrow$ reject H_0 $\rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow$ X is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support H_A $\rightarrow$ **cannot reject** H_0 $\rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow$ X is not an important predictor of y and we should not include X into our model.

Hypothesis Testing in Regression Analysis

The most often testsed hypothesis in regression analysis is:

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

It tells about the importance of the predictors included in the model:

- Conclude that there is enough evidence to support $H_A \rightarrow$ reject $H_0 \rightarrow$ there is a correlation between X and y and $\hat{\beta}$ is a statistically significant coefficient $\rightarrow X$ is an important predictor of y and we should include X into our model.
- Conclude that there is not enough evidence to support $H_A \rightarrow$ **cannot reject** $H_0 \rightarrow$ no correlation between X and y could be identified and $\hat{\beta}$ is not a statistically significant coefficient $\rightarrow X$ is not an important predictor of y and we should not include X into our model.

Decission Rule

How to decide whetherH_0 should or should not be rejected in favor of H_A?

1. The range of possible values of $\hat{\beta}$ is divided into two regions, an *"acceptance" region* and a *rejection region*
2. To define these regions, we must determine the critical values of $\hat{\beta}$ that divide the "acceptance" region from the rejection region.
3. We will reject H_0 in favor of H_A if the estimated $\hat{\beta}$ falls into the rejection region. Else we will say that we cannot reject H_0.

Decission Rule

How to decide whether H_0 should or should not be rejected in favor of H_A?

1. The range of possible values of $\hat{\beta}$ is divided into two regions, an *"acceptance" region* and a *rejection region*
2. To define these regions, we must determine the critical values of $\hat{\beta}$ that divide the "acceptance" region from the rejection region.
3. We will reject H_0 in favor of H_A if the estimated $\hat{\beta}$ falls into the rejection region. Else we will say that we cannot reject H_0.

Decission Rule

How to decide whetherH_0 should or should not be rejected in favor of H_A?

1. The range of possible values of $\hat{\beta}$ is divided into two regions, an *"acceptance" region* and a *rejection region*
2. To define these regions, we must determine the critical values of $\hat{\beta}$ that divide the "acceptance" region from the rejection region.
3. We will reject H_0 in favor of H_A if the estimated $\hat{\beta}$ falls into the rejection region. Else we will say that we cannot reject H_0.

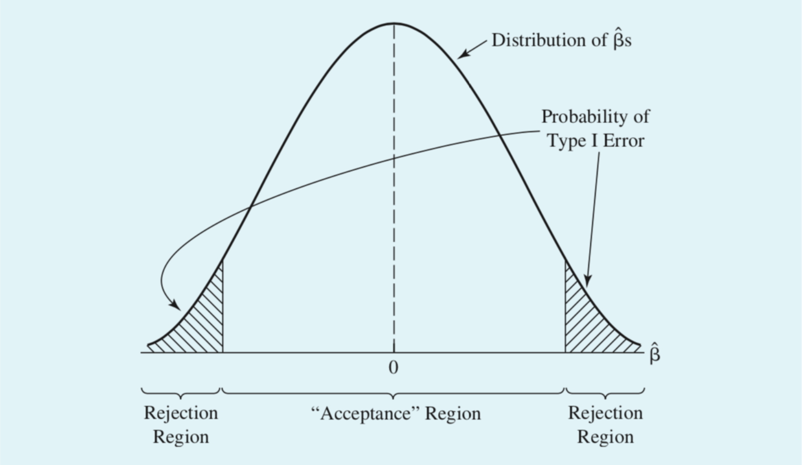
Decission Rule

How to decide whetherH_0 should or should not be rejected in favor of H_A?

1. The range of possible values of $\hat{\beta}$ is divided into two regions, an *"acceptance" region* and a *rejection region*
2. To define these regions, we must determine the critical values of $\hat{\beta}$ that divide the "acceptance" region from the rejection region.
3. We will reject H_0 in favor of H_A if the estimated $\hat{\beta}$ falls into the rejection region. Else we will say that we cannot reject H_0.

Rejection versus Acceptance Regions

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$



t-statistic

The problem with this approach is that because each X_k has a different measurement scale we will have to define a different critical value for each $\hat{\beta}_k$.

Instead we can use the *t*-statistic and the *t*-test for hypothesis testing

t-statistic

The problem with this approach is that because each X_k has a different measurement scale we will have to define a different critical value for each $\hat{\beta}_k$.

Instead we can use the <u>t-statistic</u> and the <u>t-test</u> for hypothesis testing

t-statistic

The problem with this approach is that because each X_K has a different measurement scale we will have to define a different critical value for each $\hat{\beta}_K$.

Instead we can use the *t*-statistic and the *t*-test for hypothesis testing

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

For each coefficient in the regression we can compute a t-statistic

t-statistic

The problem with this approach is that because each X_K has a different measurement scale we will have to define a different critical value for each $\hat{\beta}_K$.

Instead we can use the t-statistic and the t-test for hypothesis testing

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

For each coefficient in the regression above we can compute a t-test

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

The advantage is that because t-statistic is a standardized statistic, we will not need to find a critical value for each $\hat{\beta}_K$.

Instead we will use a predifined table to find the appropriate critical value of t_K denoted by t_c.

t-statistic

The problem with this approach is that because each X_K has a different measurement scale we will have to define a different critical value for each $\hat{\beta}_K$.

Instead we can use the t-statistic and the t-test for hypothesis testing

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

For each coefficient in the regression above we can compute a t-test

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

The advantage is that because t-statistic is a standardized statistic, we will not need to find a critical value for each $\hat{\beta}_k$.

Instead we will use a predifined table to find the appropriate critical value of t_k denoted by t_c.

t-statistic

The problem with this approach is that because each X_K has a different measurement scale we will have to define a different critical value for each $\hat{\beta}_K$.

Instead we can use the <u>t-statistic</u> and the <u>t-test</u> for hypothesis testing

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

For each coefficient in the regression above we can compute a t-test

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

The advantage is that because t-statistic is a standardized statistic, we will not need to find a critical value for each $\hat{\beta}_k$.

Instead we will use a predifined table to find the appropriate critical value of t_k denoted by t_c.

Example: Woody's Restaurant New Location

$$\hat{Y}_i = 102{,}192 - 9075N_i + 0.3547P_i + 1.288I_i$$
$$\qquad\qquad (2053) \quad (0.0727) \quad (0.543)$$
$$t = -4.42 \qquad 4.88 \qquad 2.37$$
$$N = 33 \qquad \bar{R}^2 = .579$$

Question: Is P_i an important predictor?
In other words is β_P significantly different from zero?
or just is β_P statistically significant?

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Example: Woody's Restaurant New Location

$$\hat{Y}_i = 102{,}192 - 9075N_i + 0.3547P_i + 1.288I_i$$
$$\quad (2053) \quad (0.0727) \quad (0.543)$$
$$t = -4.42 \quad 4.88 \quad 2.37$$
$$N = 33 \quad \bar{R}^2 = .579$$

Question: Is P_i an important predictor?

In other words is β_P significantly different from zero?
or just is β_P statistically significant?

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Example: Woody's Restaurant New Location

$$\hat{Y}_i = 102{,}192 - 9075N_i + 0.3547P_i + 1.288I_i$$
$$(2053) \quad (0.0727) \quad (0.543)$$
$$t = -4.42 \quad 4.88 \quad 2.37$$
$$N = 33 \quad \bar{R}^2 = .579$$

Question: Is P_i an important predictor?

In other words is β_P significantly different from zero?

or just is β_P statistically significant?

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Example: Woody's Restaurant New Location

$$
\begin{aligned}
\hat{Y}_i &= 102{,}192 - 9075N_i + 0.3547P_i + 1.288I_i \\
&\qquad\qquad (2053) \quad (0.0727) \quad (0.543) \\
&\qquad t = -4.42 \qquad 4.88 \qquad 2.37 \\
&\qquad N = 33 \qquad \bar{R}^2 = .579
\end{aligned}
$$

Question: Is P_i an important predictor?
In other words is β_P significantly different from zero?
or just is β_P statistically significant?

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Example: Woody's Restaurant New Location

$$\begin{aligned}
\hat{Y}_i = 102{,}192 - &9075N_i + 0.3547P_i + 1.288I_i \\
&(2053) \quad (0.0727) \quad (0.543) \\
t = &-4.42 \qquad 4.88 \qquad 2.37 \\
N = 33 \quad &\bar{R}^2 = .579
\end{aligned}$$

Question: Is P_i an important predictor?
In other words is β_P significantly different from zero?
or just is β_P statistically significant?

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Step 1 Hypothesis Testing: find the t-statistic

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Step 1 Hypothesis Testing: find the t-statistic

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Find the t-statistic for P:

Step 1 Hypothesis Testing: find the *t*-statistic

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Find the *t*-statistic for *P*:

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

Step 1 Hypothesis Testing: find the *t*-statistic

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Find the *t*-statistic for *P*:

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

$$t_k = \frac{(\hat{\beta}_k - 0)}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

Step 1 Hypothesis Testing: find the *t*-statistic

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Find the *t*-statistic for *P*:

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

$$t_k = \frac{(\hat{\beta}_k - 0)}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

Step 1 Hypothesis Testing: find the *t*-statistic

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Find the *t*-statistic for *P*:

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

$$t_k = \frac{(\hat{\beta}_k - 0)}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

$$\hat{Y}_i = \underset{}{102{,}192} - \underset{(2053)}{9075N_i} + \underset{(0.0727)}{0.3547P_i} + \underset{(0.543)}{1.288I_i}$$

Step 1 Hypothesis Testing: find the t-statistic

$$H_0 : \beta_P = 0$$
$$H_A : \beta_P \neq 0$$

Find the t-statistic for P:

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

$$t_k = \frac{(\hat{\beta}_k - 0)}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K)$$

$$\begin{aligned}\hat{Y}_i = 102{,}192 - &\underset{(2053)}{9075N_i} + \underset{(0.0727)}{0.3547P_i} + \underset{(0.543)}{1.288I_i}\end{aligned}$$

$$t_P = \frac{\hat{\beta}_P}{SE(\hat{\beta}_P)} = \frac{0.3547}{0.0727} = 4.88$$

Step 2 Hypothesis Testing: find the t-critical

Decide whether to reject or not to reject H_0 based on the critical t-value which distinguishes the acceptance region from the rejection region.

Let's inspect the t-table

1) Find the Degrees of Freedom: $df = N - K - 1$

$$\begin{aligned}\hat{Y}_i = 102{,}192 &- 9075N_i + 0.3547P_i + 1.288I_i \\ &\quad (2053) \quad\ (0.0727) \quad (0.543) \\ t = &\ -4.42 \qquad 4.88 \qquad\ 2.37 \\ N = 33 &\qquad \bar{R}^2 = .579\end{aligned}$$

$df = N - K - 1 = 33 - 3 - 1 = 29$

Step 2 Hypothesis Testing: find the t-critical

Decide whether to reject or not to reject H_0 based on the critical t-value which distinguishes the acceptance region from the rejection region.

Let's inspect the t-table

1) Find the Degrees of Freedom: $df = N - K - 1$

$$\hat{Y}_i = 102{,}192 - 9075N_i + 0.3547P_i + 1.288I_i$$
$$(2053) \quad (0.0727) \quad (0.543)$$
$$t = -4.42 \quad 4.88 \quad 2.37$$
$$N = 33 \quad \bar{R}^2 = .579$$

$df = N - K - 1 = 33 - 3 - 1 = 29$

Step 2 Hypothesis Testing: find the t-critical

Decide whether to reject or not to reject H_0 based on the critical t-value which distinguishes the acceptance region from the rejection region.

Let's inspect the t-table

1) Find the Degrees of Freedom: $df = N - K - 1$

$$\hat{Y}_i = 102{,}192 - 9075N_i + 0.3547P_i + 1.288I_i$$
$$\qquad\qquad (2053) \qquad (0.0727) \qquad (0.543)$$
$$t = -4.42 \qquad 4.88 \qquad 2.37$$
$$N = 33 \qquad \bar{R}^2 = .579$$

$df = N - K - 1 = 33 - 3 - 1 = 29$

Step 2 Hypothesis Testing: find the t-critical

Decide whether to reject or not to reject H_0 based on the critical t-value which distinguishes the acceptance region from the rejection region.

Let's inspect the t-table

1) Find the Degrees of Freedom: $df = N - K - 1$

$$\begin{aligned}
\hat{Y}_i &= 102{,}192 - 9075N_i + 0.3547P_i + 1.288I_i \\
& \qquad\qquad\quad (2053) \qquad (0.0727) \quad (0.543) \\
t &= \quad -4.42 \qquad\quad 4.88 \qquad\quad 2.37 \\
N &= 33 \qquad \bar{R}^2 = .579
\end{aligned}$$

$df = N - K - 1 = 33 - 3 - 1 = 29$

Step 2 Hypothesis Testing: find the t-critical

2) Decide on the **Significance Level** - α

The significance level measures the amount of Type I Error implied by a particular t-critical.

For example: If the level of significance is 10%, and we reject H_0 at 10% significance level $\rightarrow$ 10% percent of the time H_0 was correct, but we rejected it.

Think of it as a margin of error that you accept.

Step 2 Hypothesis Testing: find the t-critical

2) Decide on the **Significance Level** - α

The significance level measures the amount of Type I Error implied by a particular t-critical.

For example: If the level of significance is 10%, and we reject H_0 at 10% significance level $\rightarrow$ 10% percent of the time H_0 was correct, but we rejected it.

Think of it as a margin of error that you accept.

Step 2 Hypothesis Testing: find the t-critical

2) Decide on the **Significance Level** - α

The significance level measures the amount of Type I Error implied by a particular t-critical.

For example: If the level of significance is 10%, and we reject H_0 at 10% significance level $\rightarrow$ 10% percent of the time H_0 was correct, but we rejected it.

Think of it as a margin of error that you accept.

Step 2 Hypothesis Testing: find the t-critical

2) Decide on the **Significance Level** - α

The significance level measures the amount of Type I Error implied by a particular t-critical.

For example: If the level of significance is 10%, and we reject H_0 at 10% significance level $\rightarrow$ 10% percent of the time H_0 was correct, but we rejected it.

Think of it as a margin of error that you accept.

Step 2 Hypothesis Testing: find the t-critical

2) Decide on the **Significance Level** - α

The significance level measures the amount of Type I Error implied by a particular t-critical.

For example: If the level of significance is 10%, and we reject H_0 at 10% significance level $\rightarrow$ 10% percent of the time H_0 was correct, but we rejected it.

Think of it as a margin of error that you accept.

Step 2 Hypothesis Testing: find the t-critical

So let's say we selected the significance level $\alpha = 10\% = 0.1$

Remember that $df = 29$

Inspect the table to find t-critical

$t_c = 1.699$

Step 2 Hypothesis Testing: find the t-critical

So let's say we selected the significance level $\alpha = 10\% = 0.1$

Remember that $df = 29$

Inspect the table to find t-critical

$t_c = 1.699$

Step 2 Hypothesis Testing: find the *t*-critical

So let's say we selected the significance level $\alpha = 10\% = 0.1$

Remember that $df = 29$

Inspect the table to find *t*-critical

$t_c = 1.699$

Step 2 Hypothesis Testing: find the t-critical

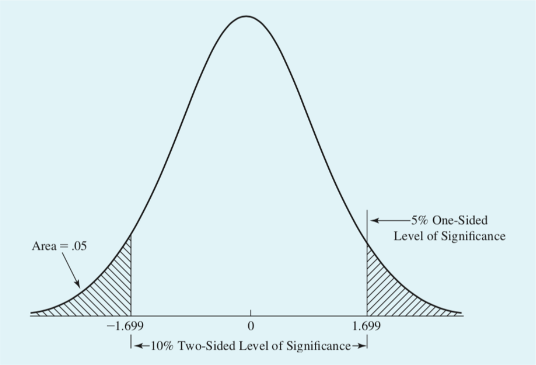
So let's say we selected the significance level $\alpha = 10\% = 0.1$

Remember that $df = 29$

Inspect the table to find t-critical

$t_c = 1.699$

Step 3 Hypothesis Testing: Decide to reject or to not reject H_0



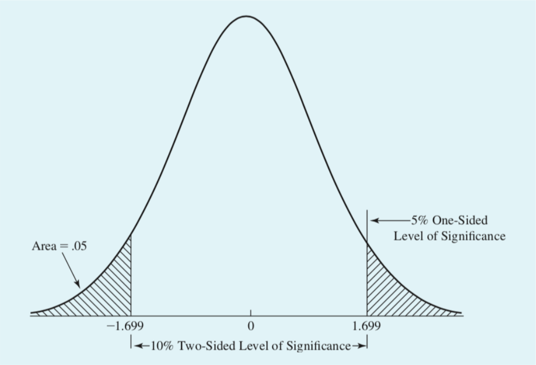
Decision Rule:

Reject H_0 if $-1.699 < t_p > 1.699$ else do not reject.

$t_p = 4.88 > 1.699 \rightarrow$ Reject $H_0 : \beta_P = 0$ in favor of $H_A : \beta_P \neq 0$

Conclussion: β_P is statistically significant and P is an importan predictor of Y

Step 3 Hypothesis Testing: Decide to reject or to not reject H_0



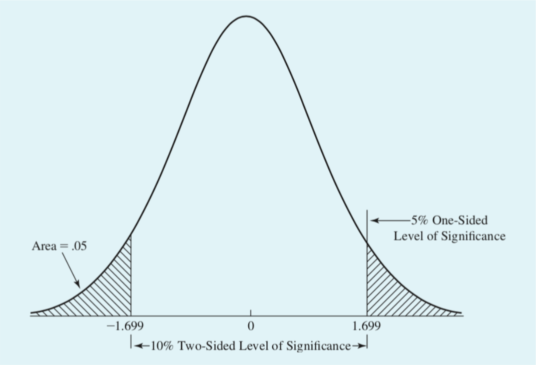
Decision Rule:

Reject H_0 if $-1.699 < t_p > 1.699$ else do not reject.

$t_p = 4.88 > 1.699 \rightarrow$ Reject $H_0 : \beta_P = 0$ in favor of $H_A : \beta_P \neq 0$

Conclussion: β_P is statistically significant and P is an importan predictor of Y

Step 3 Hypothesis Testing: Decide to reject or to not reject H_0



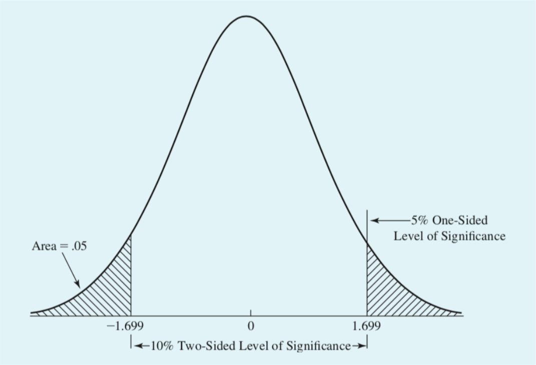
Decision Rule:

Reject H_0 if $-1.699 < t_p > 1.699$ else do not reject.

$t_p = 4.88 > 1.699 \rightarrow$ Reject $H_0 : \beta_P = 0$ in favor of $H_A : \beta_P \neq 0$

Conclussion: β_P is statistically significant and P is an importan predictor of Y

Step 3 Hypothesis Testing: Decide to reject or to not reject H_0



Decision Rule:

Reject H_0 if $-1.699 < t_p > 1.699$ else do not reject.

$t_p = 4.88 > 1.699 \rightarrow$ Reject $H_0 : \beta_P = 0$ in favor of $H_A : \beta_P \neq 0$

Conclussion: β_P is statistically significant and P is an importan predictor of Y

Let's get Started!

Access Google Colaboratory through your Gmail account